The Observer Class Hypothesis

Travis Garrett*
Perimeter Institute for Theoretical Physics, Waterloo, Ontario N2L 2Y5, Canada and
Department of Physics & Astronomy, Louisiana State University, Baton Rouge, LA 70802, USA
(Dated: January 12, 2011)

The discovery of a small cosmological constant has stimulated interest in the measure problem. One should expect to be a typical observer, but defining such a thing is difficult in the vastness of an eternally inflating universe. We propose that a crucial prerequisite is understanding why one should exist as an observer at all. We assume that the Physical Church Turing Thesis is correct and therefore all observers (and everything else that exists) can be described as different types of information. We then argue that the observers collectively form the largest class of information (where, in analogy with the Faddeev Popov procedure, we only count over "gauge invariant" forms of information). The statistical predominance of the observers is due to their ability to selectively absorb other forms of information from many different sources. In particular, it is the combinatorics that arise from this selection process which leads us to equate the observer class $\mathcal{O}$ with the nontrivial power set $\hat{\mathcal{P}}(\mathcal{U})$ of the set of all information $\mathcal{U}$. Observers themselves are thus the typical form of information. If correct, this proposal simplifies the measure problem, and leads to dramatic long term predictions.

arXiv:1101.2198v1 [gr-qc] 11 Jan 2011

I. INTRODUCTION

The detection of ongoing cosmological acceleration was one of the most interesting discoveries in recent years [1],[2]. The form of the acceleration (with w close to -1) leads to the conclusion that it is probably due to a small cosmological constant Λ [3],[4],[5]. Einstein's "biggest blunder" has turned out not to be a blunder at all.

In a further amusing twist some explanations for the source of this Λ involve arguments about typical observers within a vast multiverse. A large amount of fine tuning is needed to reduce Λ from its natural value of M_P^4, and this fine tuning winds up at a number very close to the anthropic bound, confirming Weinberg's prediction [6]. Furthermore string theory provides a wide enough landscape of states to generate such a small Λ [7–10], and eternal inflation gives a method to populate them [11],[12]. These anthropic arguments [13–17] thus reintroduce the somewhat unusual subject matter of observers in physics, some 80 years after the discovery of modern quantum mechanics in the late 1920s.

Indeed, one should expect to exist as a "typical observer": one shouldn't see 50σ events or macroscopic violations of the 2nd law and so forth. This is a reflection of the statistical nature of our universe, and practically speaking it is implicitly assumed in all scientific disciplines. The issue of what constitutes a typical observer becomes somewhat nontrivial however when considering the immense volumes of spacetime generated in cosmology.

For instance, upon making some initially reasonable-sounding assumptions one could come to the conclusion that we exist in the earliest civilization possible after the big bang (the youngness paradox [18],[19]), or that the typical observer is a Boltzmann Brain (i.e. a disembodied brain floating in the void [20],[21]). Trying to fix these bizarre results composes the measure problem: how to define probabilities in the infinite volumes generated by eternal inflation [22–37]. A proposed measure solution will posit some form of cutoff, so that only a finite number of observers are considered, with the statistical distribution of their experiences (very broadly speaking) thus well defined. There are measures that just consider the observations made along a single world line, and others that sample from entire spacetime volumes. The hope is that with the correct choice of measure the standard, non-exotic experiences of observers within our Hubble volume will be found to be typical within the entire multiverse, although this is still an open question.

There is a very deep and interesting assumption behind all of these ideas, one which is generally not questioned. As noted, one should expect to exist as a typical observer. However, why should one expect to exist as an observer at all? This is a nonintuitive subject matter (and it is something that is quite easy to take for granted), so it is helpful to take some time and rephrase the question in a number of different ways.

It is important to first clarify: what are observers? We staunchly support the materialist viewpoint and posit that the brain/mind connection is complete: all of one's conscious experiences as an observer are generated by the interactions of neurons in one's brain (see e.g. [38],[39], [40],[41]). In turn, observers exist due to natural selection: being able to extract pertinent information from one's environment is a very useful evolutionary adaptation.

To give a more detailed example, consider an individual having breakfast in the morning. At one moment in time they will experience a particular mixture of sights and sounds – the sunlight coming through the window, the coffee brewing – perhaps while having a thought about the upcoming day or remembering something from

*Electronic address: tgarrett@perimeterinstitute.ca

the previous one, and all of this is permeated by an emotional milieu and a sense of self in the background. All of these aspects of consciousness are equal to the patterns created by the subset of neurons that are active in the individual's brain at that moment in time.

To transition to a more abstract description, the experiences of that individual are a particular type of pattern: a form of topological connected graph, as realized by the connections of the currently interacting neurons. The precise structure of the pattern (and thus the nature of the conscious experiences) will change from moment to moment, based on the dynamics of the neural biophysics and the details of the incoming sensory information. There is a large amount of self-similarity however in the structure of the individual's neural patterns as they evolve in time, and indeed there is substantial similarity between different people's conscious states (e.g. we can talk to each other). Collectively all of these types of patterns form a class of mental states. To rephrase the typical observer assumption, one expects to exist as an element of this class, where each member is a type of pattern formed out of an arrangement of atoms.

Further abstraction is possible and is useful. The key is that observers are just a particular type of information, as is everything else. That is, we assume that the Physical Church Turing Thesis (PCTT) is correct [42],[43],[44], and the universe can be simulated to arbitrary precision inside a sufficiently powerful computer (e.g. in a universal Turing machine $\mathcal{T}$). This confidence is based both on the wide range of individual scientific successes to date, and the impressive way in which they overlap to form a unified whole. By calculating of all of the interactions among the constituent particles in a simulation $\mathcal{S}$ of a large region of the universe, one could thereby reproduce stellar evolution and the formation of planets, the emergence of replicating molecular structures, and after a long period of Darwinian evolution the arrival of complex life forms which can observe and understand their environment.

If all of the details in the history of events inside a particular simulation $\mathcal{S}_i$ happen to match the evolution within our Hubble volume, then the observers within that simulation would be indistinguishable from us: the patterns generated by their simulated neurons would precisely match the patterns formed in our brains. At the same time, and at the deep level of the hardware generating the simulation, those observers are just another type of information: complex sequences of 1s and 0s that evolve according to the implemented algorithm. Everything else in that simulation, or in any other running program, is also just information – different sequences of binary data.

Describing everything that exists in terms vast numbers of logical operations is admittedly a strident form of reductionism, and is of little practical use in most day to day activities, scientific or otherwise. One primary purpose of this gedankenexperiment is to help shake off the natural tendency to take existing as an observer for granted: those observers in the giant simulation are completely equivalent to the information encoded in sequences of 1s and 0s (as is everything else that exists). What is the special property of the observer-sequences so that one should exist as one of them, rather than some other form of information?

Before proposing an answer, let us first examine the nature of information in a bit more detail. Note that the precise form of the various sequences of 1s and 0s is not necessarily critical. For instance, a Turing machine $\mathcal{T}_1$ running the simulation $\mathcal{S}_i$ can itself be emulated by a second universal Turing machine $\mathcal{T}_2$. The various objects inside the simulation $\mathcal{S}_i$ (say, comets and seagulls and picnics and so forth) will be encoded in different binary sequences in $\mathcal{T}_1$ and $\mathcal{T}_2$, but the real information content of these objects does not change.

There is thus a degree of redundancy possible when describing objects in a computational framework (or any other framework). Consider the set $\mathcal{A}$ given by the union of the output of all possible programs. $\mathcal{A}$ can be generated by acting on all finite length binary strings s_i with a universal Turing machine $\mathcal{T}$:

$$\mathcal{A} = \bigcup \mathcal{T}(s_i). \tag{1}$$

There will then be an equivalence class $[\mathcal{S}_i]$ in $\mathcal{A}$ for all of the programs that encode the same history of events described by $\mathcal{S}_i$:

$$[\mathcal{S}_i] = \{\mathcal{S}_j \in \mathcal{A} | \mathcal{S}_i \sim \mathcal{S}_j\}. \tag{2}$$

More broadly, the elements of $\mathcal{A}$ can be partitioned into the "information structure" blocks $[x_i]$:

$$\mathcal{A} \,/ \sim_{en} = \{[x_i] | x_i \in \mathcal{A}\} \tag{3}$$

where we use the special equivalence relation $\sim_{en}$ which has the abstract definition: "encodes the same information". For instance, in addition to detecting and grouping together all the programs that give rise to the simulation $\mathcal{S}_i$, this equivalence relation will also group together the data streams of the programs that give a proof of the Pythagorean theorem, or describe the life cycle of a species of fern, and so forth. We will later discuss the feasibility of explicitly constructing such a general equivalence relation.

Alternatively, a particular binary representation x_{a1} for some information structure x_a encoded on $\mathcal{T}_1$ amounts effectively to a gauge choice for the description of that structure – similar to choosing a coordinate system g_{ab} to map a spacetime manifold $\mathcal{M}$. Furthermore, we will argue, in an analogy with the Faddeev Popov procedure ([45],[46]), that it is useful to get rid of the redundant gauge descriptions when counting among various forms of information. We will therefore borrow the term "gauge" (in a slight abuse of the term) to describe the various representations that are possible for any information structure. We are primarily interested in the translationally invariant information x_a, with various gauge choices $G(x_a)$ being more or less helpful in describing it.

We have assumed that the PCTT is correct, and thus everything that exists can be represented as various types of information x_i (for convenience we will drop the brackets from $[x_i]$). We will name the union of all these structures the universe of all information $\mathcal{U}$:

$$\mathcal{U} = \bigcup x_i = \mathcal{A} \,/\, \sim_{en} \,. \tag{4}$$

$\mathcal{U}$ resembles the von Neumann universe V of set theory (which is a proper class [47]), or Tegmark's Level 4 Multiverse [48],[49]. We adopt a Platonic viewpoint and assume that the entirety of $\mathcal{U}$ actually exists. $\mathcal{U}$ thus contains mathematical structures like $x_{3p} = \{3 \in P\}$ (where P are the primes), $x_{eul} = \{e^{i\pi} + 1 = 0\}$, the Pythagorean theorem x_{pt}, the Mandelbrot set x_{ms}, the wave equation x_{we} and so on.

$\mathcal{U}$ also contains very complex structures like our physical universe, which we will denote by Ψ (i.e. $\Psi \subset \mathcal{U}$). Through objects like Ψ the universal set will also contain all of the complex emergent phenomena that can arise through the interactions of many particles, such as volcanos x_{vol}, grasshoppers x_{gh}, constitutional monarchies x_{cm}, and sensory qualia x_{sq}. These emergent structures are somewhat statistical in nature, as defining them will involve some degree of coarse graining over microscopic degrees of freedom.

Observers are included among these complex structures, and we will grant them the special name y_j (although they are also another variety of information structure x_i). For instance a young child y_{c1} may know about x_{3p} and x_{gh}: $x_{3p}, x_{gh} \in y_{c1}$, while having not yet learned about x_{eul} or x_{cm}. This is the key feature of the observers that we will utilize: the y_j are entities that can absorb various x_i from different regions of $\mathcal{U}$.

We can now rephrase our central question in its final form. Consider the Universal Set $\mathcal{U}$: the set of all forms of information. All of the observers y_j will collectively form a particular subset of the Universal Set: the Observer Class $\mathcal{O}$:

$$\mathcal{O} = \bigcup y_j. \tag{5}$$

It is a given that to exist at all entails existing as some form of information – to be some element within $\mathcal{U}$. Why is it that one exists as an element of the Observer Class $\mathcal{O}$ in particular? This question is shown pictorially in Fig. 1, where the different forms of information are symbolically represented as various strings of 1s and 0s (i.e. some computational gauge has been chosen).

The Observer Class Hypothesis (OCH) proposes a statistical answer: observers form by far the largest subset of $\mathcal{U}$:

$$|\mathcal{O}| \gg |\mathcal{U} - \mathcal{O}| \tag{6}$$

and thus a element chosen randomly from the universal set is overwhelmingly likely be an observer (i.e. Fig. 1 is not drawn to scale). Just as typical observers dominate the counting among all observers, observers themselves dominate the counting among all forms of information.

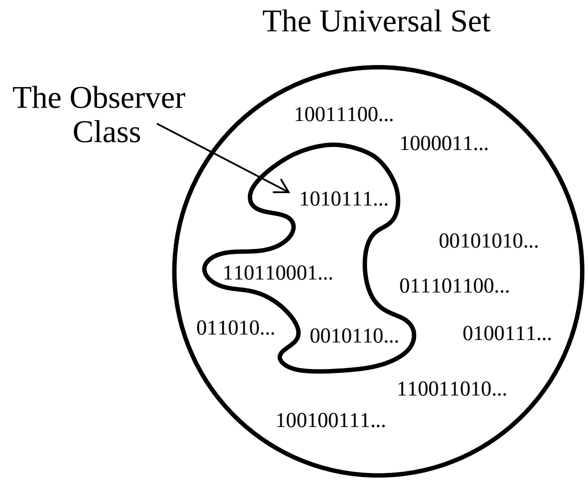

FIG. 1: The Observer class as a subset of the Universal Set (represented in some computational gauge).

The observers form the most numerous class of information due to their primary trait: they can observe other forms of information, and thus incorporate them as subsets of themselves. For N different pieces of information x_i in $\mathcal{U}$, there will generally be 2^N different possible observers y_j who absorb anywhere from 0 to all N of the x_i. The observers thus collectively resemble the power set $\mathcal{P}(\mathcal{U})$ of the set of all information, and indeed we will later propose specifically that they effectively form the nontrivial power set $\hat{\mathcal{P}}(\mathcal{U})$:

$$\mathcal{O} \sim \hat{\mathcal{P}}(\mathcal{U}). \tag{7}$$

Note also that observers can observe other observers (and so on recursively) and so for example $\mathcal{P}(\mathcal{P}(\mathcal{U})) \subset \mathcal{O}$. We therefore refer to $\mathcal{O}$ as a class (i.e. a proper class) rather than a set. In summation it is the combinatorial potential of the observers which leads them to vastly outnumber all other types of information, and thus provide the general form that existence invariably takes.

II. GAUGES AND COUNTING

The core idea of the Observer Class Hypothesis is that if one counts over all forms of information they will find that observers greatly outnumber all other structures. However, the counting process seems to be problematic at first for a number of reasons. For example, there is the issue of over-counting pieces of information that have merely been re-expressed in a different form. It would also appear that completely random structures in fact dominate the counting instead of observers. These issues are solvable, which can be demonstrated through an examination of gauge choice (i.e. the selection of some formalism to represent different types of information). The concept of gauge invariance then leads to a more precise definition of information, and influences the considera-

tion of the number of different forms of information that exist.

We first examine the gauge redundancy that exists in the representation of all forms of information. Consider as an example the small element x_{a1} inside of $\mathcal{U}$: $x_{a1} = \{3 \in P\}$, where P are the prime numbers. This simple mathematical factoid x_a can be re-expressed in an infinite number of different ways: $x_{a2} = \{1 + 2 \in P\}$, $x_{a3} = \{\sqrt{9} \in P\}$, and so on (i.e. $x_{a1}, x_{a2}, x_{a3} \in G(x_a)$). The proposal thus appears to be plagued by infinities, as even one simple mathematical statement would get counted an infinite number of times.

This over counting problem is not limited to the mathematical equivalence class. To pick a literary example, one can take a famous quote from Hamlet, and arbitrarily translate it into the "1-q" language: "qto qbe qor qnot qto qbe", or into the "2-q" language and so forth. In general this form of vacuous translational redundancy exists for all information structures. This can be seen in the example we noted before, where a simulation $\mathcal{S}_i$ running on the Turing machine $\mathcal{T}_1$ can be emulated by a second universal Turing machine $\mathcal{T}_2$. The history of events within $\mathcal{S}_i$ won't change, but the sequence of 1s and 0s describing it in $\mathcal{T}_2$ does.

It also appears at first that the universe $\mathcal{U}$ is in fact dominated by random noise. If we continue with the computational framework and consider the elements of $\mathcal{U}$ to be binary strings, then there are 2^N different strings s_i that are N bits long. The vast majority of these 2^N strings are completely random, and do not encode any real gauge-invariant information. This is reflected, for instance, in the fact that the Kolmogorov complexity $K(s_i)$ [50],[51] of an average string s_i is comparable to the length of s_i:

$$K(s_i) \sim |s_i| \tag{8}$$

We will denote a true random string (i.e. it has no internal patterns or real structure) by r_i. This is distinct from the Kolmogorov definition of randomness, which equates being incompressible with being random. Most of the incompressible s_i strings are indeed patternless r_i strings, but a nontrivial fraction will instead be the most compact representations of real information structures (e.g. consider the most compact binary description of the Einstein equations x_{EE} versus a random string r_i of the same length).

We propose that both of these problems – over-counting due to gauge redundancies and the apparent numerical dominance of the random noise structures – can be solved by choosing a gauge. Consider a similar problem that arises in quantum field theory due to gauge fields. A straightforward evaluation of the path integral Z:

$$Z = \int DA e^{\frac{i}{\hbar}\int \mathcal{L}(A,\partial A)d^4x} \tag{9}$$

for a gauge field A_μ gives an infinite result since the Lagrangian $\mathcal{L}$ is invariant under gauge transformations $A_\mu \to A'_\mu = A_\mu + \partial_\mu\phi$. Faddeev and Popov give us a solution: separate out the redundant gauge portion of the Lagrangian and integrate it separately as an irrelevant infinite constant. Colloquially speaking, we get the correct answer if we only consider the physical variations of A_μ, and not the redundant gauge variations $G(A_\mu)$.

We posit that the same thing is true when counting over all forms of information: we need to count over only the real, distinct information structures (such as $x_{a1} = \{3 \in P\}$), and ignore the redundant translations (such as $x_{a2} = \{2 + 1 \in P\}$ and so forth) which would lead to an infinite over-counting. The particular representation formalism chosen to describe the different types of information is not important here, as only its ability to remove the extraneous redefinitions is needed.

We also postulate that the selection of a representation gauge removes all of the random noise structures r_i which appear at first to dominate the counting in $\mathcal{U}$. In short, the random structures are all gauge translations of nothing at all. Consider for instance the quote from Hamlet, which we can arbitrarily translate into the "318-q" language, which makes the quote much longer but doesn't add any real information to it. In the same fashion, a very long (and internally structureless) random sequence r_i is nothing more than the r_i translation of the null set $\emptyset$:

$$r_i \in G(\emptyset) \tag{10}$$

The intuitive feeling that random structures don't contain any real, nontrivial information is indeed correct, and a manifestation of this is that they do not carry any gauge-invariant information (see also [52],[53]).

In contrast, non-trivial forms of information contain internal patterns and structures that map unambiguously from representation to representation: a child will learn that a mouse is a type of mammal in an English, French, or Chinese biology book. In comparison, one is free to translate a book-length random sequence r_i into any other random sequence they wish, as it does not have any internal structures that must be preserved. We thus conclude that the true information content of the book $I(s_{bio})$ is much larger than the information content of a random string of the same length : $I(s_{bio}) \gg I(r_i) \sim 0$. This is opposed to the result one gets from the Kolmogorov measure, where $I_K(s_{bio}) < I_K(r_i)$ since the biology book will be compressible to a degree, while the random sequence r_i is not. Focusing on the gauge invariant information is therefore equal to selecting objects that have nontrivial "effective complexity" [54] or a large amount of "logical depth" [55],[56]. As the random and redundant elements are removed by the selection of a gauge, we can then use the Kolmogorov measure to describe the complexity of the various x_i (e.g. $I_K(s_{bio})$ for the biology book). There will then be many more x_i with large values for I_K than those with a small amount of complexity, and thus a randomly chosen information structure will likely be very complex.

We have described the real, non-trivial information

structures as having gauge invariant internal patterns, but have not given an explicit characterization of this property. To this end one imagines a finite length algorithm F that can search through all 2^N of the N bit strings s_i (for any N) and extract and organize all of the gauge invariant information structures x_i, with the remainder being the gauge translations of those structures $G(x_i)$, or the random strings $r_i = G(\emptyset)$. For instance F would be able to examine the strings $s_p = \{3 \in P\}$ and $s_q = \{4 \in P\}$, and select s_p to be filed among the x_i, while discarding s_q, or determine that the effective complexity of a frog is much greater than a glass of water of the same mass.

As a concrete example, consider the simple world of binary addition ψ_{BA}, which is composed of strings of 4 symbols: 0,1,+,=. In general it is possible to have correct strings: "1+1=10", incorrect strings: "1+0=11", and ill-formed strings: "+1+==". For this toy model one can construct an algorithm F_{BA} that will search through all 4^N of the N symbol strings and extract the fraction which are correct. A numerical implementation of F_{BA} shows that the number of correct strings scales approximately as $\sim 0.17 \times 2.53^N$. This is the type of behavior that we expect in the general case: the number of real structures will grow exponentially (via combinatorics) as the string length N increases, but at a slower rate than the total number of strings. The x_i will thus form a sparse subset of the strings s_i, with a density trending towards zero as $N \to \infty$. On a side note, we do not conclude that there is an infinite amount of real information in ψ_{BA}, even though there are correct strings of unbounded length. Rather, the amount of nontrivial information in ψ_{BA} is comparable to the length of the algorithm F_{BA}: $I(\psi_{BA}) \sim |F_{BA}|$.

One can construct these types of algorithms for various simple problems, but it is not possible to construct a finite length algorithm F for the general case, as demonstrated by Gödel's Incompleteness Theorems [57], and Church and Turing's negative answer to Hilbert's Entscheidungsproblem [58],[59]. This itself is not a problem, but rather a reflection of the natural structure of $\mathcal{U}$. Just as random strings r_i can't be compressed, the union of real information structures (4) can not be completely encapsulated by any finite length description. Ignoring categorization, there is not even a universal trait that all of the individual x_i carry which a "litmus test" type F could test for: no single equation in Wiles' papers proves Fermat's Last Theorem, no single chemical bond in a chlorophyll molecule shows that photosynthesis is a useful evolutionary adaptation. One needs essentially a library (or the internet) to give a zeroth order approximation of F for merely the things we have learned so far in our corner of Ψ. The best one can do for a concise measure of information content is thus something like the Kolmogorov measure, which works due to the pigeonhole principle, but is necessarily silent on the semantic content of the various binary strings.

Explicitly constructing a finite algorithm F that sifts through all binary strings s_i and extracts all of the x_i which compose $\mathcal{U}$ is not possible, but in the end it isn't needed either. To make an analogy, Ψ does not actually physically perform a Faddeev Popov type procedure to remove the redundant gauge degrees of freedom every time two gluons interact (which alternative formulations like BRST show). Alternatively, all of the various programs $\mathcal{T}(s_i)$ in $\mathcal{A}$ will either halt or not, there just isn't a finite length algorithm that can determine which is the case in general. Likewise, $\mathcal{U}$ simply has a natural, intrinsic structure, which the consideration of gauge invariance merely helps to highlight. This natural structure is clearest in the case of the mathematical x_i: it really is the case that 3 is a prime number and 4 is not (see also [60],[61]). Just because we consider the union of all x_i, it is emphatically not the case that "anything goes".

The existence of this natural structure also carries over to the complex emergent phenomena that can arise in physical universes like Ψ. For instance, a salamander (x_{sal}) is a nontrivial information structure which occurs in our universe. One consequence of being a real x_i is that salamanders exist much more frequently in our Hubble volume than a direct counting of all the permutations Ω of the same number of atoms would suggest. In fact, we will later use a high frequency of occurrence $f(x_i)$ as an indicator that an emergent object x_i has nontrivial structure.

The intricacy of these complex emergent phenomena can be seen if we further prod our example structure x_{sal}: that is, what do we mean precisely by a salamander? A certain individual organism at one moment in time? Do we coarse grain over the orientations of the water molecules in that individual's cells? What about the particulars of the ongoing gene expression in those cells at that moment in time? Do we instead consider all the individuals in a breeding population at a moment in time, or the general features of their shared genotype and phenotype over a broader span? Furthermore there has been a nearly continuous evolution (being discrete at the parent-child level) from the most basic replicating molecular structures early in the Earth's history to the various species that exist today (see, e.g. [62]) – where does one make a cutoff?

Naturally, there is no single right answer, as different definitions and levels of detail illuminate different aspects of the information structure which we concisely named x_{sal}. While very complex and inextricably interwoven with many other objects (and local to the branch of Ψ we find ourselves in), x_{sal} is also not arbitrary or random: it is an example of a natural structure occurring in $\mathcal{U}$, not unlike the mathematical structures $x_a = \{3 \in P\}$ or $x_b = \{e^{i\pi} + 1 = 0\}$. Note, however, that in fact all possible arrangements of atoms will occur with at least some non-zero frequency inside an eternally inflating Ψ. Just existing as a pattern somewhere within a physical universe like Ψ is thus too imprecise of a measure for membership in $\mathcal{U}$. Rather, we generally want the patterns that occur frequently in Ψ to be promoted to the

x_i elements of $\mathcal{U}$ – those patterns that minimize the Lagrangian in (9), roughly speaking.

In fact, there is a natural solution for selectively ranking the various complex patterns and phenomena that can arise in a universe like Ψ: do not consider the patterns in perfect isolation (like pure mathematical factoids), but rather as the emergent structures embedded in a physical universe that they are. The frequency f at which a particular pattern occurs in Ψ (which is closely linked to the process that generates it) thus becomes a useful tool in measuring its information content. For instance, all sorts of bizarre objects (say, interplanetary china teapots: x_{ict}) can be formed in an eternally inflating Ψ through macroscopic quantum tunneling events. As these objects are formed through abrupt random processes they contain little gauge-invariant information, and this in turn is linked to their exponentially small production rates (which also corresponds to having very little logical depth [55]). In short, they can be quickly and concisely described as very improbable random fluctuations, and thus for example: $I(x_{sal}) \gg I(x_{ict})$.

Note that the frequency at which a particular pattern x_a occurs (which is a function of x_a's internal structure) is only one component of its information content – there is also the actual complexity $\sim I_K(x_a)$ of the pattern (at an appropriately coarse-grained level). The gauge invariant information contained in a pattern x_a will thus roughly scale as the complexity of the pattern times the frequency of its occurrence:

$$I(x_a) \propto I_K(x_a) \times f(x_a). \tag{11}$$

$I_K(x_a)$ is the baseline Kolmogorov complexity of the pattern, and a large value for $f(x_a)$ then makes it likely that the value of $I_K(x_a)$ stems from nontrivial effective complexity rather than random noise (e.g. the compact description of the Einstein equations x_{EE} will have a large value for f as compared to a vanishing value for a random sequence r_i of the same length).

In fact, it would generally suffice for our purposes to define f to be a step function, so that it equals 1 for those patterns that have a greater than 50 percent chance of occurring within a Hubble volume over the 100 billion years of standard stellar evolution, and zero otherwise. In this way f does the work of F and effectively extracts the sparse subset of nontrivial structures from among the exponentially large set of possibilities. For instance, if we coarse grain biological patterns at the cellular level (with, say, 100 different cell types, each with an average mass of 10^{-9} grams), then there are $\sim 100^{10^{14}}$ different possible arrangements for a 100 kilogram "organism". In a similar fashion to the simple binary addition example ψ_{BA}, the number of viable organisms (i.e. $f = 1$) will grow exponentially with increasing mass, but at a much slower rate than the total number of arrangements (i.e most permutations have $f = 0$).

A step function profile for f is admittedly artificial – in reality f will be a function of the internal structure of the various patterns (and the environment in which they occur). In general, there is a smooth continuum from exceptionally unlikely to commonplace phenomena, and thus there is no clean cutoff such that we could include the more common objects to be among the real information structures x_i while discarding the rest as random r_i. Indeed, evolution via natural selection, lacking goals or foresight, must sift through many different patterns (of varying degrees of randomness) in a process of iterative trial and error in order to generate nontrivial emergent structures such as x_{sal}. Likewise, even with goals and foresight, iterative trial and error is an indispensable component of scientific discovery – as observed for instance by Poincaré. We are thus not calling for the elimination all aspects of randomness, but rather merely positing that the apparent preponderance of the completely random r_i in $\mathcal{U}$ was always an illusion.

By including the frequency at which various complex emergent phenomena occur as a useful indicator of their information content we essentially rediscover the measure problem. Within the confines of our Hubble volume things are clear cut: spontaneous violations of the 2nd law are very rare (say, a glass of water separating out into ice and steam), and thus have a very low information content. In the expanse of an infinite universe however, even these very rare events can occur an infinite number of times, and the counting thus becomes trickier. It would seem that with a countably infinite number of both the "rare" and "common" events that one is free to arrange and count them in any fashion and thus arrive at any relative frequency. A common example of this is to ask what fraction of the integers are even. A natural guess is 1/2: $\{1, 2, 3, 4, 5, 6, \dots\}$, but they can be rearranged to give any fraction between zero and one – say 1/3: $\{1, 3, 2, 5, 7, 4, \dots\}$.

In response to this we ask a related question: what fraction of the integers are prime? The same trick can be played here, with the primes and composites being rearranged to give any desired fraction. However, if we make the restriction of considering all integers less than N, and then letting N go to infinity, then there is a nontrivial answer: the number of primes $\pi(N)$ less than or equal to N is approximately equal to $N/\ln(N)$ (or, more closely, $\mathrm{Li}(N) = \int_2^N dt/\ln(t)$). If the Riemann hypothesis is correct we can even add the order of magnitude error estimate: $\pi(N) = \mathrm{Li}(N) + O(\sqrt{N}\ln(N))$. Just asking what fraction of the integers are prime is thus an incomplete question. The question is uniquely finished by adding the constraint (or regularization) of considering all integers less than N (and then letting $N \to \infty$). Furthermore, the correctness of this particular question can then be seen by the abundance of natural structure (the Prime Number Theorem x_{pnt}, and the associated properties of the Riemann Zeta function) that it leads to.

We assert that the same scheme holds for physical universes like Ψ: they can be infinite in extent and still contain nonrandom, nontrivial structure (which can be revealed by asking the right questions). Agricultural so-

cieties, for instance, will occur much more frequently in Ψ type universes than teapots spontaneously formed in protoplanetary disks. This is certainly the case within the confines of our Hubble volume, and the puzzle is to then find the correct regularization or measure that extends this general pattern of relative frequencies to arbitrarily larger volumes. Note that it is not the case that the relative frequencies in distant regions will be identical to those we observe locally – but rather it should emerge that our Hubble volume is typical among the subset of regions that support complex life. This problem is a good bit more subtle than the case of regulating the integers, and we will not offer our own measure proposal here. We will also refrain from endorsing other measure proposals, although we think there are interesting candidates that may be on the right track.

Our primary goal is even broader in scope than recovering the general distribution of probabilities within a physical universe like Ψ. We claim that it is not only the case that the x_{sal} type objects greatly outnumber the random x_{ict} objects in a Ψ type universe (i.e. we assume that a correct regularization which confirms one's intuition is possible), but that the x_{sal} outnumber the x_{ict} throughout the entirety of $\mathcal{U}$ (and therefore the nonrandom structures dominate the information content of $\mathcal{U}$).

To give a possible counterexample to this proposal, there will be "heat bath" type universes ψ_{HB} within $\mathcal{U}$, such that all of their internal patterns are generated through random fluctuations. In these ψ_{HB} universes the x_{sal} and x_{ict} type structures will be produced at comparable rates. Since no consistent patterns emerge, the interactions that occur within a spacetime volume of one of these ψ_{HB} universes can effectively be compressed to the initial state and evolution equations that then evolve it. Ψ on the other hand starts with a very low initial entropy (instead of a maximal entropy like the ψ_{HB}), and will thus preferentially generate certain complex emergent patterns (the "viable ones") as the entropy increases (and these emergent structures are not described by the initial conditions and evolution equations, other than in some "latent" sense). The heat bath universes are thus effectively the analogue of the long random strings r_i: they appear very complex at first, but they carry little real information – essentially only that carried in their evolution equations. The information content $I(\Psi)$ within the Ψ type universes is thus vastly greater than the random ψ_{HB} type universes: $I(\Psi) \gg I(\psi_{HB})$.

In addition to the completely disjoint ψ_{HB} type universes existing elsewhere within $\mathcal{U}$, it is possible that our own universe could evolve into an eternal heat bath phase. This would be the case, for instance, if the current acceleration is due to a true cosmological constant Λ, so that we will eventually transition to an eternal de Sitter universe, with a horizon temperature of: $T_{dS} = \sqrt{2G\rho_\Lambda/3\pi}$. A straightforward counting of all of the observers y_j in this scenario would seem to be dominated by the infinite number of Boltzmann Brain observers y_{BB} in the cold de Sitter phase. In turn this is sometimes presented as evidence that Λ must not stable and will decay at some point in the future. There are independent reasons to be skeptical of an eternal de Sitter phase [63], and we are not proposing that our universe has a true, immutable Λ. However, we do not think the possibility of an infinite number of y_{BB} rules a true Λ out: the real information is contained in the initial, stellar evolution phase of Ψ.

Finally, one can also conjecture the existence of other types of physical universes (i.e. Ψ^{-1}) where the internal dynamics (during an entropy increasing phase) specifically leads to the bizarre objects like the interplanetary teapots consistently outnumbering the natural objects like salamanders. If these universes were to exist, then upon counting over both the Ψ and Ψ^{-1} one could find that all possible patterns are produced at comparable rates. Here we simply assert, given the inherent structure of $\mathcal{U}$ and the deep mathematics that provide the foundation of Ψ, that these Ψ^{-1} type universes do not exist – just as the statements $\{e^{i\pi} + 1 = 4\}$, or "the Mandelbrot algorithm produces an isosceles triangle" do not correspond to x_i in $\mathcal{U}$. The nonrandom emergent patterns that arise in nontrivial mathematical structures like Ψ are therefore also the primary content of $\mathcal{U}$ itself.

In summation, the universe of all information is full of nontrivial mathematical structures and complex emergent phenomena – i.e. the sorts of things that observers are interested in absorbing. Indeed, it is because to this ability to selectively extract information from other sources that observers themselves form the vast majority of $\mathcal{U}$.

III. OBSERVERS

The Observer Class Hypothesis proposes that the observers y_j collectively form the largest subset of the universe of all information $\mathcal{U}$. In particular, the way in which the observers selectively absorb x_i from various regions of $\mathcal{U}$ (say, $x_a, x_b, x_d, ... \in y_\alpha; x_a, x_c, x_e, ... \in y_\beta; ...$), results in the observer class resembling the power set $\mathcal{P}(\mathcal{U})$. We will refine this observation, and propose specifically that observers effectively form the nontrivial power set of $\mathcal{U}$. The "non-triviality" restriction for the elements of $\mathcal{P}(\mathcal{U})$ then naturally leads to physically realized observers existing within a superstructure like Ψ. We then consider the overall size of $\mathcal{U}$, including the possibility that it is infinite in extent. If $I(\mathcal{U}) = \aleph_0$, then universal observers $\hat{y}_j$ may be able to absorb any x_i by continuously upgrading their ability to process information.

Consider first the straightforward power set $\mathcal{P}(\mathcal{X})$ of some set of information $\mathcal{X} = \bigcup_{i=1}^{N} x_i \subset \mathcal{U}$. The amount of information in $\mathcal{X}$ can be found by summing the information content of its members:

$$I(\mathcal{X}) = \sum_{i=1}^{N} I_K(x_i) \tag{12}$$

(where one can use the Kolmogorov measure I_K as the alternative gauge versions $G(x_i)$ and random strings r_i have already been cut from $\mathcal{U}$). What then is the information content of $\mathcal{P}(\mathcal{X})$? Each element $x_a \in \mathcal{X}$ will occur in half of the 2^N elements of $\mathcal{P}(\mathcal{X})$, so a direct counting would give $I_{dir}(\mathcal{P}(\mathcal{X})) = 2^{N-1}I(\mathcal{X})$. This is a very inefficient description however – the entire content of $I(\mathcal{P}(\mathcal{X}))$ is in fact contained in the combination of the original set, and in the definition of the power set, so that the true information content is: $I(\mathcal{P}(\mathcal{X})) = I(\mathcal{X}) + I(\mathcal{P}) \sim I(\mathcal{X})$.

However, it is also possible to combine basic information structures and form a new, nontrivial structures through their interconnections. For instance, various aspects of Riemannian geometry ($x \in \mathcal{X}_{Riem}$) were used in the derivation of the Einstein equations of general relativity, or in Penrose and Hawking's singularity theorems, and in Perelman's proof of the Poincaré conjecture. Different combinations of proteins ($x \in \mathcal{X}_{prot}$) allow assorted bacteria to live in environments ranging from the root systems of fig trees, to the shores of alpine lakes or in thermal vents on the sea floor. Alternatively, various arrangements of electrical components ($x \in \mathcal{X}_{elec}$) allow for the creation of radios, radars and integrated circuits.

We therefore define the nontrivial power set $\hat{\mathcal{P}}(\mathcal{X})$, to be the sparse subset of combinations from $\mathcal{X}$ that lead to new structures in $\mathcal{U}$:

$$\hat{\mathcal{P}}(\mathcal{X}) = \mathcal{P}(\mathcal{X}) \bigcap \mathcal{U}. \tag{13}$$

There are a couple caveats in this definition. In general the repetition of simple elements in $\mathcal{X}$ is allowed (e.g. many transistors are needed for a computer chip). Additionally the elements of $\mathcal{X}$ may need a system of tags so that the interconnections within the new structures can be explicitly identified (say, the locations of the circuit elements on the computer chip). The straightforward power set of some set $\mathcal{X}$ (of N elements) does not add any new information, but the selection of the nontrivial combinations does:

$$I(\mathcal{X}) \sim I(\mathcal{P}(\mathcal{X})) \ll I(\hat{\mathcal{P}}(\mathcal{X})). \tag{14}$$

The elements of $\hat{\mathcal{P}}(\mathcal{X}) - \mathcal{X}$ are new x_i elements of $\mathcal{U}$ (which observers can also absorb).

Consider then the proposed observer y_{r1} (i.e. a direct element of $\mathcal{P}(\mathcal{U})$): $y_{r1} = \{x_{tang}, x_3, x_{nept}\}$, where x_{tang} is a tangerine, x_3 is the number 3, and x_{nept} is the planet Neptune. This random collection of various information structures from $\mathcal{U}$ is clearly not an observer, or any other from of nontrivial information: y_{r1} is redundant to its three elements, and would thus be cut by the selection of a gauge. This is the sense in which most of the direct elements of the power set of $\mathcal{U}$ do not add any new real information.

However, one could have a real observer y_α whose main interests happened to include types of fruit, the integers, and the planets of the solar system and so forth. The 3 elements of y_{r1} exist as a simple list, with no overarching structure actually uniting them. A physically realized computer, with some finite amount of memory and a capacity to receive input, resolves this by providing a unified architecture for the nontrivial embedding of various forms of information. A physical computer thus provides the glue to combine, say, x_{tang}, x_3, and x_{nept} and form a new nontrivial structure in $\mathcal{U}$.

It is possible to also consider the existence of "randomly organized computers" which indiscriminately embed arbitrary elements of $\mathcal{U}$ – these too would conform to no real x_i. This leads to the specification of "physically realized" computers, as the restrictions that arise from existing within a mathematical structure like Ψ results in computers that process information in nontrivial ways. Furthermore, a structure like Ψ allows for these physical computers to spontaneously arise as it evolves forward from an initial state of low entropy. Namely it is possible for replicating molecular structures to emerge, and Darwinian evolution can then drive to them to higher levels of complexity as they compete for limited resources. A fundamental type of evolutionary adaptation then becomes possible: the ability to extract pertinent information from one's environment so that it can be acted upon to one's advantage. The requirement that one extracts useful information is thus one of the key constraints that has guided the evolution of the sensory organs and nervous systems of the species in the animal kingdom.

This evolutionary process has reached its current apogee with our species, as our brains are capable of extracting information not just from our immediate surroundings, but also from more abstract sources such as $\mathcal{X}_{Riem}$, $\mathcal{X}_{prot}$, or $\mathcal{X}_{elec}$. We can absorb the thoughts of others on wide ranging subjects, from Socrates' ethics, to Newton's System of the World. Observers can also create structure, from individual works like van Gogh's paintings, to collective entities like the global financial system. The combinatorics that arise from the expansive scope of sources that Homo Sapiens can extract information from thus explains why we are currently the typical observers.

We have hypothesized that observers comprise the largest class of information, but the implications that follow from this depend on the size of $\mathcal{U}$ itself. In general there are three possibilities for the extent of the universe of all information:

- $\mathcal{U}$ is finite: $I(\mathcal{U}) < N$ (for some finite N),
- $\mathcal{U}$ is countably infinite: $I(\mathcal{U}) = \aleph_0$,
- $\mathcal{U}$ is uncountably infinite: $I(\mathcal{U}) > \aleph_0$.

Roughly speaking, $\mathcal{U}$ could be "small", "big", or "very big". It could be argued that the first case is the null assumption, but we will also seriously examine the possibility that $I(\mathcal{U}) = \aleph_0$ (which follows if there is no finite upper bound to the complexity of real information structures). However, we will not inspect in depth the $I(\mathcal{U}) > \aleph_0$ case (which would necessitate x_i that are irreducibly infinitely complex). It is conceivable that the third case could be meaningful if hypercomputers of some

form could be built, but we find this scenario unlikely. We note in passing that if was the case that $I(\mathcal{U}) > \aleph_0$, then it may be possible to demonstrate (6) via cardinality arguments.

First consider the case that $\mathcal{U}$ is finite. One begins with an infinite number of programs producing data in $\mathcal{A}$, but the action of removing both the redundant translations and the random output significantly prunes the binary tree. It is possible that after this gauge cut only a finite number of finitely complex structures x_i remain, so that the sum of all their information content is also a finite number. Note that this is possible since we use compact representations for the various x_i. For instance, one could go through and list all of the prime integers, thereby generating an infinite number of true statements, but the existence of an infinite number of primes is concisely expressed in Euclid's proof. Likewise, a single uncomputable real requires an infinitely long description, but the general procedure to generate reals can be concisely defined by Dedekind cuts (therefore most real numbers are effectively infinitely long r_i sequences). In this sense there may only be a finite number of unique mathematical structures, encapsulated in some $\mathcal{X}_{math}$, so that no combinations from $\mathcal{P}(\mathcal{X}_{math})$ lead to new, nontrivial structures.

Furthermore, if $I(\mathcal{U}) < N$, then our physical universe Ψ only contains a finite amount of real information. Ψ could still be infinite in spatial extent, perhaps due to a form of eternal inflation. However, in this scenario, the dynamics of Ψ would only allow finitely complex objects to arise in any one region. For instance, the deepest dynamics of Ψ may be completely described by some finitely complex mathematical structure (perhaps a completion of string theory). These fundamental dynamical laws may only admit a finite number of degrees of freedom in any finite volume of spacetime (in agreement with current Planck scale/Holography proposals [64],[65],[66]). The construction of arbitrarily complex structures would then necessitate coherent communication across arbitrarily large scales. If it is the case that $I(\mathcal{U}) < N$, then unbounded communication of this sort is not possible. Additionally, if $\mathcal{U}$ is finite, then necessarily there are not arbitrarily more complex alternative universes ψ elsewhere in $\mathcal{U}$ – the complexity of the structures within our Ψ would be representative for the finite number of other physical universes in $\mathcal{U}$.

If the Observer Class Hypothesis is correct and $\mathcal{U}$ is finite, then in principle a direct counting would show $|\mathcal{O}| \gg |\mathcal{U} - \mathcal{O}|$. In practical terms there would be physical observers y_j with the memory capacity to absorb a large number of the most complex x_i in $\mathcal{U}$ – combinatorics would then ensure that the observers dominate the counting. The number of x_i structures with a Kolmogorov complexity of N bits should then roughly look like a rescaled and truncated Poisson distribution $f(N;\lambda)$, with the observers y_j composing the majority of the central bulk (and with a typical observer describable by $\sim \lambda$ bits).

This is the "small" case for $\mathcal{U}$, with a finite upper bound for the complexity I_K of all gauge invariant information structures when expressed in a compact representation (although this "small" case still corresponds to a very rich and intricate universe of information). If this is the nature of $\mathcal{U}$ then presumably as typical nontrivial observers we should already be asymptotically approaching everything that can be known. Of course, it is also possible for $\mathcal{U}$ to be finite but for the OCH to be incorrect. For instance there could be a very large $\mathcal{Z} \subset \mathcal{U}$ such that all of the elements of $\mathcal{Z}$ are much too complex to be absorbed by any physical observer (and thus $|\mathcal{Z}| \gg |\mathcal{O}|$). In this case there must be some special property possessed by the observers, other than being very numerous, so that one should exist as an observer, rather than be an element of $\mathcal{Z}$ (note that we still assume the PCTT and thus everything that exists is some form of information). Note that as we do happen to be observers we are effectively "trapped": we can only point to the possible existence of a vast $\mathcal{Z}$. Any progress towards explicating its various structures would drive $\mathcal{Z}$ to being a subset of $\mathcal{O}$. In general we will make the simplifying assumption that these enormous and "forever inaccessible" $\mathcal{Z}$ sets do not exist.

It would be interesting if $I(\mathcal{U}) < N$, but the alternative is even more intriguing: it may well be that $\mathcal{U}$ is composed of an infinite number of distinct, finitely complex x_i. We view Gödel's Incompleteness result (paraphrased as "No finite system of axioms and inference rules captures all mathematical truths"), and Turing's halting result ("no finite algorithm can capture the behavior of all programs") as suggestive that $I(\mathcal{U}) = \aleph_0$. Likewise, the probable result from computational complexity that P $\neq$ NP (and the associated complexity class hierarchy [67]) is also suggestive that there are nontrivial objects of unbounded complexity in $\mathcal{U}$. However, we do not interpret these results as definitive proof that $\mathcal{U}$ is infinite. For instance, repeatedly applying Gödel's result to generate new independent axioms doesn't actually lead to new structure, and the unpredictability of most programs may be due their being effectively random in structure.

If the OCH is correct, and $I(\mathcal{U}) = \aleph_0$, then the status of observers is more subtle. As there are x_i of unbounded complexity in $\mathcal{U}$ there would need to be arbitrarily complex y_j to absorb those structures. If these y_j do exist the counting would need to be regulated by first considering only x_i with a Kolmogorov complexity of less than N bits, and then letting $N \to \infty$. If the OCH holds, then the density of the observers $|\mathcal{O}|/|\mathcal{U}|$ will approach one as the complexity of structures goes to infinity. In this way the observers would roughly resemble the composite integers, with the primes corresponding to the various information structures that they can absorb.

This proposal appears at first to be incompatible with our existence as a typical observers. For instance, consider a class $Y_\alpha \subset \mathcal{U}$ which is composed of observers one thousand times more complex than a Homo Sapiens: $I_K(y_a) \sim 10^3 I_K(y_b)$, with $y_a \in Y_\alpha$ and $y_b \in Y_{HS}$. Via

combinatorics it follows that $|Y_\alpha| \gg |Y_{HS}|$ and thus one should expect to be one of the very advanced Y_α instead of a human. However, this process can be repeated indefinitely: one should instead expect to be a member of Y_β whose members are 10^3 times more complex than those in Y_α and so on. As there is no upper bound, there is no one level of complexity that one should expect to exist at – there will always be many more individuals at much higher levels.

We propose that the resolution for this is that one should expect to exist as a universal observer $\hat{y}$, rather than existing at any one level of complexity. A universal observer $\hat{y}$ has the ability to absorb any information structure x_i from $\mathcal{U}$, and they thus collectively dominate the counting among all observers (as well as $\mathcal{U}$ itself). However, in order to be nontrivial information structures we have argued that the observers must be concretely realized computers embedded in some mathematical structure like Ψ. Being concretely realized, an observer will have a finite capacity at any one moment in time, and accordingly there will be an upper bound to the complexity of objects that they can promptly absorb at that time. The solution is to not place a time limit for the absorption of a very complex structure: in general a universal observer $\hat{y}_j$ will need to gradually upgrade their physical capacity (via scientific and technological progress) for a long period of time until they have the ability to absorb an arbitrarily large x_i from $\mathcal{U}$. Universal observers are thus self similar: at any one moment in time they will find themselves to be more complex than some other individuals, and less complex than a large number of (potential) observers above them. Note in particular that they are not structures that exist statically at any one point in time, but rather are entities that are always evolving in time. It is because of this capacity for growth that the universal observers – both concretely realized and continuously evolving – are the typical observers (if $I(\mathcal{U}) = \aleph_0$).

This scenario seems to be at odds with our current experiences: after reaching adulthood the complexity of what one can learn generally levels off. If this state of affairs was to persist indefinitely, then one should conclude something along the lines of: $\mathcal{U}$ is finite and we are close to saturating what is possible, or: $\mathcal{U}$ is infinite in extent and therefore the OCH is wrong. However, there is a real chance that this restriction could be lifted in the coming decades. In particular the development of human level artificial intelligence (and then beyond) would open the possibility of unbounded growth. We are not yet universal observers, but we may be on the boundary of becoming them.

Admittedly, the plausibility of the development of human level Artificial Intelligence (AI) is a contentious issue, and there have been promises for its arrival at dates that have long since come and gone [68]. Still, it can be argued that computing hardware has now become sufficiently advanced to support strong AI. For instance, based on analysis of the retina Moravec [69] estimates the information processing rate of a single neuron to be on the order 10^3 flops, which results in an estimate of 10^{14} flops for the entire brain, which is comparable to the current largest supercomputers. Likewise there is steady progress on the software side in a large number of specialized fronts. To give a concrete example, given the advances in robotics, natural language processing, and object recognition, it should now be possible (with sufficient funding) to build a bipedal robot which, if asked, could find and describe various objects while navigating its environment. There is still a large distance between this type of robot and an AI that could, say, be a good sitcom writer, but this gulf should be bridgeable by continuous evolutionary improvements. The bridging might not even take too long: Kurzweil, for instance, optimistically predicts that strong AI capable of passing the Turing test could arrive by as early as 2030 [70]. His predictions are based on a coarse grained extrapolation of the growth in information technologies, which so far have followed smooth exponentials for many decades.

In general, most non-dualists would agree that human level AI is possible in theory, if perhaps disagreeing on a likely arrival date. Note also that while the emergence of AI is not assured in any one civilization, that it is effectively guaranteed to occur in some fraction of histories in an eternally inflating universe. Furthermore, if it is developed, it would be surprising if it were to then stall permanently at about our intelligence level (although this could be evidence that we are already saturating what is possible in a finite $\mathcal{U}$). It seems more likely that they would be able keep on progressing, in time processing thoughts orders of magnitude more complex than we are capable of now. However, there currently appear to be limits to the complexity of computers that could conceivably be built, which puts the prospect of unbounded growth in doubt [71],[72]. As noted, Planck scale and holographic arguments indicate that a finite number of degrees of freedom are available for computation in a finite volume of spacetime, and the construction of arbitrarily large structures would at least be a nontrivial undertaking.

These apparent roadblocks are not necessarily insurmountable. For instance, Dyson [73] argued that life could persist indefinitely in a flat or open universe (if in an ever slower, colder form). Krauss and Starkman disagree [74], pointing out that with a nonzero Λ that temperatures can not be lowered indefinitely (although note [63]), and alternatively in a flat or open $\Lambda = 0$ universe the amount of matter available to a civilization would either be finite, or would lead to a collapse into a black hole. However, they observe that their analysis does not examine speculative strong quantum gravity effects (such as [75]) which may allow workarounds to these restrictions (we also note that true singularities may be avoided in the interiors of black holes - see e.g. [76],[77]).

Likewise, the apparent restrictions on the number of degrees of freedom available in a finite volume are not necessarily inviolable. These restrictions stem from the current structure of proposed fundamental theories, but

these may evolve and generalize in the future. To pick an example, the majority of calculations in string theory have been done in the critical dimension $D = 10$, but non-critical (linear dilaton) theories are also conceivable [78],[79]. More abstractly, it has always been the case in the past that the theories we have discovered have been the limiting cases (e.g. $\hbar \to 0; v/c \to 0$) of more complex theories, and this may apply to string theory (and thus its implications) as well. Furthermore, if it is possible for an apparently fundamental theory to be a particular limit of a yet more complex structure, then statistically this should turn out to be the case (i.e. there will be many "deeper structure" variations available in $\mathcal{U}$, as opposed to only one version where the deepest laws are equal to those that have already been discovered).

Finally, the development of powerful AI seems possible within standard atomic physics, and so the subsequent details in the task of unbounded advancement would fall to them (assuming they do arrive). Note that a degree of vagueness in these futuristic predictions is unavoidable due to the sparseness of the nontrivial structures x_i among all possible permutations s_i (compounded by the likely fact that P$\neq$NP). Scientific and technological advancement necessarily involves a large amount of trial and error because of this sparseness, and is thus both "hard" and hard to predict. Still, if $\mathcal{U}$ is infinite and the OCH is correct, then in time the AIs will be successful in finding and utilizing modifications of (or loopholes in) basic physics in order to continuously compute and absorb evermore complex information structures, thus becoming universal observers $\hat{y}_j$.

IV. APPLICATION AND PREDICTIONS

In his "Discourse on the Method" René Descartes searched for the securest foundation for philosophical inquiry, and converged upon "I think, therefore I am". We disagree with his subsequent claims, but the central observation is quite interesting. The insight can be split into two parts: "I exist" (or just: "Existence") and then "I exist as a thinker". We assume basic existence (i.e. that there is something rather than nothing), and question the second part: why is it that to exist, to be anything at all, entails existing as a thinker? Why is it that to be an observer is the particular form that existence takes?

We have proposed a statistical answer in the OCH. The first clue emerges from the great unity in the world that the scientific revolution has revealed in the last 350 years. While many of the details remain to be filled in, the broad sweep of phenomena that we observe fits neatly into an overarching explanatory structure, from mathematics and particle physics to atomic physics and chemistry, up through microbiology and neurology to consciousness and the evolution of civilization. It is striking that all of the objects in this expansive range can be fully described as various types of information (assuming the physical version of the Church Turing Thesis is correct). The universal capacity of information for representation is likewise reflected in the way in which all the elements of conscious experience (from any type of idea to memories, emotions, sensory qualia, and the sense of self) are equal to the different patterns created by interacting neurons.

Following this observation it is natural to postulate a "grand unification": everything that exists is a type of information. The x_i can be collected in a $\mathcal{U}$, which removes gauge redundancies and random strings. The observers then form a subset of the grand ensemble with the special property that they selectively extract information from other sources. If the Observer Class Hypothesis is correct, the combinatorics that arise from this selection process lead to the observers dominating the counting among all information structures (6). This provides a statistical explanation for why we find ourselves to be thinkers: it is by far the most probable form of existence.

We have considered two specializations of the hypothesis: in one version $I(\mathcal{U}) < N$ there is a finite upper bound to the complexity of objects x_i (when expressed in a compact representation), and in the other $I(\mathcal{U}) = \aleph_0$ there is no upper bound. The first possibility could be viewed as the null assumption – currently the complexity of the universe looks finite. However, we will adopt the more extravagant second version in order to apply the OCH and make predictions. This is version is at least conceivable: for instance both Gödel and Turing's results are suggestive that there is no upper bound to the complexity of real information structures. If $I(\mathcal{U}) = \aleph_0$ then the OCH requires the existence of universal observers $\hat{y}_j$ which can absorb any information structure. Being physically realized, they will in general need to upgrade their capacity for a long period of time before they have the ability to absorb a particular x_i. The universal observers are thus concretely realized, continuously evolving in time, and broadly self-similar. The curious fact that strong AI could be developed in the coming decades is supportive of the possibility that universal observers could be developed.

We can now apply the universal observer version of the OCH to several problems involving typical observers. We first examine the Doomsday argument due to Carter [80] and Gott [81]. In the simplified version one considers 2 possible Future Histories: a "utopia" FH1, and a "doomsday" version FH2. In FH1 our species perseveres for a very long time (say $\sim 10^8$ years), so that perhaps $\sim 10^{16}$ humans will live at some point. In the alternative FH2, some terrible event eliminates our species in the fairly near future (say within the 21st century) so that the total number of humans that will ever live is comparable to the number that have been born so far (i.e. from 10^{10} to 10^{11}). Note that we are typical observers in the doomsday history FH2, but are very special observers in the FH1 scenario: in the utopian version we live very close to the beginning of history. If the prior probabilities for FH1 and FH2 are comparable (a large assumption), then using Bayesian statistics one would conclude that

the disastrous FH2 scenario will likely take place.

The conclusion is flipped if we assume that we are universal observers however. This is because the observations of $\hat{y}_j$ are roughly self similar: at any one moment in time they can promptly absorb x_i of some finite complexity, but at arbitrarily distant points in the future they will be able to absorb arbitrarily more complex structures. All universal observers therefore find themselves to be relatively close to the beginning of history.

We can also address the peculiar Boltzmann Brains (BBs) which seem to affect the choice of measure in cosmology. We argue that the assumptions of $I(\mathcal{U}) = \aleph_0$ and universal observers are not needed here: in a finite $\mathcal{U}$ one also should expect to not be a BB, even if they are generated late in the evolution of Ψ type universes. The real information content in Ψ is contained in the early non-equilibrium period where a certain subset of patterns are produced much more frequently (because of the details of their internal structure) than a direct counting of all permutations Ω would indicate. These nontrivial patterns form the information content of Ψ, and they are the objects that one could exist as. The Boltzmann Brains are random fluctuations that happen to reproduce the patterns that exist in the stellar evolution phase – they carry no information content independent of the nontrivial observers that they copy. If one could exist as a thermal fluctuation, then one would be a completely random arrangement, not one of the BBs which comprise a tiny fraction of the permutation space. Thermal fluctuations and random structures in general contain little real information (despite first appearances) as they can be efficiently encapsulated as trivial r_i (i.e. they have very low effective complexity [54]). Note that if universal observers are possible then the case against BBs is further strengthened: one should exist as a constantly evolving and growing observer, while the BBs cease to exist only moments after they are formed.

The concept of Boltzmann Brains has recently been used as a tool in the construction of measures in cosmology, and to argue that vacuum energy must decay on a fairly fast timescale. We in effect argue that this particular tool is not valid – one should exist as a real, gauge-invariant information structure, such as the ordinary observers that can emerge after several billion years of evolution. Anthropic arguments and measure proposals may still be useful, but they need to be predicated on these nontrivial observers. This is the case for instance in Weinberg's prediction for the cosmological constant, or in Dicke [82] and Carter's [83] explanations of Dirac's large number coincidences.

The OCH can also be used to make coarse-grained predictions about the evolution of minds and civilization going into the future. Previously we have considered the possibility of the development of human level AI (and then beyond) as supportive of the large $\mathcal{U}$ version of the OCH. We switch here and assume $I(\mathcal{U}) = \aleph_0$, and thus the typical observer (and therefore the typical form of information) is a universal observer $\hat{y}_j$. This then necessitates the development of strong AI, to be followed by ceaseless advancement and exploration of ever more complex x_i. Note that the universal observer version of the OCH probably can not be definitively proven as it hinges upon the existence of arbitrarily complex nontrivial structures. Rather, in practical terms, the longer that scientific and technological progress persist the more confidence can be had that it is correct. The current prediction is thus limited to the arrival of strong AI, as the detailed prospects for subsequent progress would then fall to them. A precise date for the emergence of strong AI can not be given, but it is striking that the timescale appears to be only on the order of decades (which compares favorably, for instance, to the $\sim 10^{10}$ year vacuum decay predictions).

Finally we emphasize the explanatory power of the Observer Class Hypothesis. It is easy to take existing as an observer for granted, but upon reflection there should be something special about minds so that to exist at all is to exist as one of them. The OCH answers in both ways: observers are not special as they are just another type of information, but they are special since their ability to absorb other forms of information makes them very numerous. Further consideration points towards only counting over distinct, gauge-invariant forms of information. One thus needs nontrivial observers, and these can emerge from biological evolution working within the constraints imposed by a mathematical structure like Ψ. If there is no upper bound to the complexity of information structures, then existing as a entity that is continuously evolving in time is also natural: this solves the dual constraints of being physically realized while making possible the absorption of any x_i. The most common form of information, and most probable form of existence, is thus a typical observer who is embedded in a structure like Ψ and is continuously evolving in time.

[1] A. G. Riess *et al.* [Supernova Search Team Collaboration], "Observational evidence from supernovae for an accelerating universe and a cosmological constant," Astron. J. **116**, 1009-1038 (1998). [astro-ph/9805201].

[2] S. Perlmutter *et al.* [Supernova Cosmology Project Collaboration], "Measurements of Omega and Lambda from 42 high redshift supernovae," Astrophys. J. **517**, 565-586 (1999). [astro-ph/9812133].

[3] S. Weinberg, "The cosmological constant problem," Rev. Mod. Phys. **61**, 1 (1989).

[4] M. Trodden and S. M. Carroll, "TASI lectures: Introduction to cosmology," arXiv:astro-ph/0401547.

[5] S. Nobbenhuis, "Categorizing Different Approaches to the Cosmological Constant Problem," Found. Phys. **36**, 613 (2006) [arXiv:gr-qc/0411093].

[6] S. Weinberg, "Anthropic Bound on the Cosmological Constant," Phys. Rev. Lett. **59**, 2607 (1987).

[7] R. Bousso, J. Polchinski, "Quantization of four form fluxes and dynamical neutralization of the cosmological constant," JHEP **0006**, 006 (2000). [hep-th/0004134].

[8] J. L. Feng, J. March-Russell, S. Sethi and F. Wilczek, "Saltatory relaxation of the cosmological constant," Nucl. Phys. B **602**, 307 (2001) [arXiv:hep-th/0005276].

[9] S. Kachru, R. Kallosh, A. D. Linde and S. P. Trivedi, "De Sitter vacua in string theory," Phys. Rev. D **68**, 046005 (2003) [arXiv:hep-th/0301240].

[10] L. Susskind, "The anthropic landscape of string theory," arXiv:hep-th/0302219.

[11] A. D. Linde, "Eternally Existing Selfreproducing Chaotic Inflationary Universe," Phys. Lett. **B175**, 395-400 (1986).

[12] A. Aguirre, "Eternal Inflation, past and future," arXiv:0712.0571 [hep-th].

[13] A. Vilenkin, "Predictions from quantum cosmology," Phys. Rev. Lett. **74**, 846 (1995) [arXiv:gr-qc/9406010].

[14] M. Tegmark and M. J. Rees, Astrophys. J. **499**, 526 (1998) [arXiv:astro-ph/9709058].

[15] H. Martel, P. R. Shapiro and S. Weinberg, "Likely Values of the Cosmological Constant," Astrophys. J. **492**, 29 (1998) [arXiv:astro-ph/9701099].

[16] J. Garriga and A. Vilenkin, "On likely values of the cosmological constant," Phys. Rev. D **61**, 083502 (2000) [arXiv:astro-ph/9908115].

[17] B. Carter, "Anthropic principle in cosmology," arXiv:gr-qc/0606117.

[18] A. H. Guth, "Eternal inflation and its implications," J. Phys. A **40**, 6811 (2007) [arXiv:hep-th/0702178].

[19] R. Bousso, B. Freivogel and I. S. Yang, "Boltzmann babies in the proper time measure," Phys. Rev. D **77**, 103514 (2008) [arXiv:0712.3324 [hep-th]].

[20] A. D. Linde, "Sinks in the Landscape, Boltzmann Brains, and the Cosmological Constant Problem," JCAP **0701**, 022 (2007). [hep-th/0611043].

[21] A. De Simone, A. H. Guth, A. D. Linde *et al.*, "Boltzmann brains and the scale-factor cutoff measure of the multiverse," Phys. Rev. **D82**, 063520 (2010). [arXiv:0808.3778 [hep-th]].

[22] S. Weinberg, "The cosmological constant problems," arXiv:astro-ph/0005265.

[23] S. Weinberg, "Living in the multiverse," arXiv:hep-th/0511037.

[24] M. Tegmark, A. Aguirre, M. Rees and F. Wilczek, "Dimensionless constants, cosmology and other dark matters," Phys. Rev. D **73**, 023505 (2006) [arXiv:astro-ph/0511774].

[25] B. Freivogel, M. Kleban, M. Rodriguez Martinez and L. Susskind, "Observational consequences of a landscape," JHEP **0603**, 039 (2006) [arXiv:hep-th/0505232].

[26] T. Banks and M. Johnson, "Regulating eternal inflation," arXiv:hep-th/0512141.

[27] A. Aguirre, T. Banks and M. Johnson, "Regulating eternal inflation. II: The great divide," JHEP **0608**, 065 (2006) [arXiv:hep-th/0603107].

[28] J. Polchinski, "The cosmological constant and the string landscape," arXiv:hep-th/0603249.

[29] L. Susskind, "The Census Taker's Hat," arXiv:0710.1129 [hep-th].

[30] F. Wilczek, "Enlightenment, knowledge, ignorance, temptation," [hep-ph/0512187].

[31] D. N. Page, "Return of the Boltzmann brains," Phys. Rev. D **78**, 063536 (2008) [arXiv:hep-th/0611158].

[32] J. B. Hartle and M. Srednicki, "Are We Typical?," Phys. Rev. D **75**, 123523 (2007) [arXiv:0704.2630 [hep-th]].

[33] M. Srednicki and J. Hartle, "Science in a Very Large Universe," Phys. Rev. D **81**, 123524 (2010) [arXiv:0906.0042 [hep-th]].

[34] D. N. Page, "Typicality Derived," Phys. Rev. D **78**, 023514 (2008) [arXiv:0804.3592 [hep-th]].

[35] A. Linde, M. Noorbala, "Measure Problem for Eternal and Non-Eternal Inflation," [arXiv:1006.2170 [hep-th]].

[36] A. Linde, V. Vanchurin, "How many universes are in the multiverse?," Phys. Rev. **D81**, 083525 (2010). [arXiv:0910.1589 [hep-th]].

[37] A. Aguirre, M. Tegmark and D. Layzer, "Born in an Infinite Universe: a Cosmological Interpretation of Quantum Mechanics," arXiv:1008.1066 [quant-ph].

[38] C. Koch "The Quest for Consciousness: A Neurobiological Approach," Roberts and Company Publishers, 1 ed. (2004).

[39] E. B. Baum "What is Thought?," The MIT Press, 1 ed. (2006).

[40] J. Hawkins "On Intelligence," St. Matrin's Griffon, 1 ed. (2005).

[41] A. Damasio "The Feeling of What Happens," Mariner Books, 1 ed. (2000).

[42] S. A. Kleene "Introduction to Metamathematics" Ishi Press (2009).

[43] D. Deutsch, "Quantum Theory, the Church-Turing Principle and the Universal Quantum Computer," Proceedings of the Royal Society, Series A, 400, 97-117 (1985).

[44] R. P. Geroch and J. B. Hartle, "Computability and physical theories," Found. Phys. **16**, 533 (1986).

[45] L. D. Faddeev, V. N. Popov, "Feynman Diagrams for the Yang-Mills Field," Phys. Lett. **B25**, 29-30 (1967).

[46] A. Zee, "Quantum field theory in a nutshell," Princeton, UK: Princeton Univ. Pr. (2010) 576 p.

[47] T. Jech, "Set Theory," Springer, 3rd ed. (2006).

[48] M. Tegmark, "Parallel universes," arXiv:astro-ph/0302131.

[49] M. Tegmark, "The Mathematical Universe," Found. Phys. **38**, 101 (2008) [arXiv:0704.0646 [gr-qc]].

[50] R. Solomonoff, "A Preliminary Report on a General Theory of Inductive Inference," Report V-131, Zator Co., Cambridge, Ma. Feb 4, 1960, revision, Nov., 1960.

[51] A. N. Kolmogorov, "Three approaches to the quantitative definition of information," Problems Inform. Transmission 1 (1) (1965) 1-7

[52] G. J. Chaitin, "Information randomness and incompleteness," Series in Computer Science, Papers on Algorithmic Information Theory, Singapore: World Scientific, 1990,

2nd ed.

[53] T. Tao, "The dichotomy between structure and randomness, arithmetic progressions, and the primes," [arXiv:math/0512114].

[54] M. Gell-Mann, and S. Lloyd, "Information measures, effective complexity, and total information," Complexity, 2: 44-52 (1996).

[55] C. H. Bennett, "Logical Depth and Physical Complexity," in The Universal Turing Machine a Half-Century Survey, edited by Rolf Herken, Oxford University Press, pp. 227-257 (1988).

[56] N. Ay, M. Mueller, and A. Szkola, "Effective Complexity and its Relation to Logical Depth", IEEE Trans. Inf. Th., Vol. 56/9 pp. 4593-4607 (2010) [arXiv:0810.5663 [cs.IT]].

[57] E. Nagel, J. Newman, and D. R. Hofstadter "Gödel's Proof," NYU Press, 50th Ann. Ed. (2008).

[58] A. Church "An unsolvable problem of elementary number theory," American Journal of Mathematics, 58 (1936).

[59] A. Turing "On computable numbers, with an application to the Entscheidungsproblem," Proceedings of the London Mathematical Society, Series 2, 42 (1937).

[60] T. Gowers, "The two cultures of mathematics," in: Mathematics: Frontiers and Perspectives, International Mathematical Union. V. Arnold, M. Atiyah, P. Lax, B. Mazur, Editors. American Mathematical Society, 2000.

[61] T. Tao, "What is good mathematics?," [arXiv:math/0702396].

[62] R. Dawkins, "The Ancestor's Tale: A Pilgrimage to the Dawn of Evolution," Mariner Books (2005).

[63] L. Dyson, M. Kleban and L. Susskind, "Disturbing Implications of a Cosmological Constant," JHEP **0210**, 011 (2002) [arXiv:hep-th/0208013].

[64] G. 't Hooft, "Dimensional reduction in quantum gravity," arXiv:gr-qc/9310026.

[65] L. Susskind, "The World As A Hologram," J. Math. Phys. **36**, 6377 (1995) [arXiv:hep-th/9409089].

[66] J. Polchinski, "Introduction to Gauge/Gravity Duality," arXiv:1010.6134 [hep-th].

[67] S. Aaronson, "Limits on Efficient Computation in the Physical World," arXiv:quant-ph/0412143v2.

[68] S. Russell and P. Norvig, "Artificial Intelligence: A Modern Approach," Prentice Hall, 2nd ed. (2002).

[69] H. Moravec, "Robot: Mere Machine to Transcendent Mind," Oxford University Press (2000).

[70] R. Kurzweil, "The Singularity Is Near: When Humans Transcend Biology," Penguin, 1st ed. (2006).

[71] S. Lloyd, "Computational capacity of the universe," Phys. Rev. Lett. **88**, 237901 (2002) [arXiv:quant-ph/0110141].

[72] L. M. Krauss and G. D. Starkman, "Universal Limits on Computation," arXiv:astro-ph/0404510.

[73] F. J. Dyson, "Time without end: Physics and biology in an open universe," Rev. Mod. Phys. **51**, 447 (1979).

[74] L. M. Krauss and G. D. Starkman, "Life, The Universe, and Nothing: Life and Death in an Ever-Expanding Universe," Astrophys. J. **531**, 22 (2000) [arXiv:astro-ph/9902189].

[75] A. Borde, M. Trodden and T. Vachaspati, "Creation and structure of baby universes in monopole collisions," Phys. Rev. D **59**, 043513 (1999) [arXiv:gr-qc/9808069].

[76] S. D. Mathur, "The fuzzball proposal for black holes: An elementary review," Fortsch. Phys. **53**, 793 (2005) [arXiv:hep-th/0502050].

[77] K. Skenderis and M. Taylor, "The fuzzball proposal for black holes," Phys. Rept. **467**, 117 (2008) [arXiv:0804.0552 [hep-th]].

[78] E. Silverstein, "(A)dS backgrounds from asymmetric orientifolds," arXiv:hep-th/0106209.

[79] A. Maloney, E. Silverstein and A. Strominger, "De Sitter space in noncritical string theory," arXiv:hep-th/0205316.

[80] B. Carter, "The anthropic principle and its implications for biological evolution, Phil. Trans. Roy. Soc. A310 (1983).

[81] J. R. Gott, III, "Implications of the Copernican principle for our future prospects," Nature 363 (1993).

[82] R. H. Dicke, "Dirac's Cosmology and Mach's Principle," Nature 192: (1961).

[83] B. Carter, "Large Number Coincidences and the Anthropic Principle in Cosmology," IAU Symposium 63: Confrontation of Cosmological Theories with Observational Data. Dordrecht: Reidel (1974).